In-Person Practice Midterm

Data Sets

Some questions refer to an NFL 2019 Passing data set which has a row for each NFL player and game in which the player threw a pass during weeks 1–6 of the 2019 NFL season.

Some questions refer to the Police Incident Report data set where each row is a different police incident.

The actual in-person exam will be somewhat longer than this practice exam.

Multiple Choice. Each problem is worth 3 points.

Circle the correct answer.

Problem 1. Which `ggplot2` command adds a straight regression line to a plot with a shaded uncertainty ribbon around the line? (Assume the `x` and `y` aesthetics have been set to quantitative variables.)

(a) `geom_line()`

(b) `geom_line(se = FALSE)`

(c) `geom_smooth(se = FALSE)`

(d) `geom_smooth(method = "lm")`

Problem 2. What is the term used for random displacement to the position of plotted points?

(a) `alpha` (b) `facet` (c) jitter (d) `shape`

Problem 3. Which name is **not valid** for an object in R?

(a) `calc_se` (b) `calcSE` (c) `calc-SE` (d) `calc.SE`

Problem 4. Which mutating join function keeps all rows from both data frames?

(a) `full_join()` (b) `inner_join()` (c) `left_join()` (d) `right_join()`

Problem 5. The NFL passing data set contains columns `Pos`, which contains the position, `Tm`, which is a three-letter code for the NFL of the player, and `Opp`, which is the three-letter code of the opponent team for that game. Which `filter()` command will keep all cases where the player is a quarterback (position has `QB` value) and either the team or the opponent is the Green Bay Packers (three-letter code `GNB`)?

(a) `filter(Pos = "QB" & (Tm = "GNB" | Opp = "GNB"))`

(b) `filter(Pos == "QB" & (Tm == "GNB" | Opp == "GNB"))`

(c) `filter(Pos = "QB" && (Tm = "GNB" || Opp = "GNB"))`

(d) `filter(Pos == "QB" && (Tm == "GNB" || Opp == "GNB"))`

Short Answer. Each problem is worth 5 points

Problem 6. Briefly explain when to use `geom_bar()` and when to use `geom_col()` when making bar graphs in `ggplot2`?

Problem 7. The NFL passing data set `nfl` has a column named `Player` which contains strings with the player name, a backslash character `\`, and a unique player code. Here is an example of the first three lines.

```
Lamar Jackson\JackLa00
Dak Prescott\PresDa01
Robert Griffin\GrifRoe01
```

Briefly explain how `sep="\\\\"` matches the `\` character in this code example.

```
nfl %>%
  separate(Player, into= c("Name","NameID"), sep="\\\\")
```

Problem 8. The police incident data set `police` has a column `IncidentDate` which contains the date and time of each incident. The column `IncidentType` has categorical information about each incident. One such type is "Robbery". Describe what information will be in the plot created by the following code by filling in the blanks.

```
police %>%
  mutate(wday = wday(IncidentDate, label=TRUE)) %>%
  filter(IncidentType == "Robbery") %>%
  ggplot(aes(x=wday)) +
    geom_bar()
```

The bar plot has _____ bars.
The x-axis labels begin with _____ and ends with _____.
The count for each bar represents the number of _____.

Data Analysis. Each problem is worth 15 points.

Problem 9. Each question is based on the following block of code and the following information about the NFL passing data set.

- The variable `Week` is a number from 1 to 6 and indicates in which week of the NFL season the player threw one or more passes.
- The variable `Age` has the form `years.days` and is the age of the player on the date of the game in the variable `Date`.
- The variable `Player` has been separated into `Name` and `Code`.

```
nfl9 = nfl %>%
  ## A
  filter(Week == 1) %>%
  select(Name,Age,Date) %>%
  ## convert=TRUE means the new variables are numbers and not strings
  separate(Age, into = c("age_years","age_days"),
           convert=TRUE) %>%
  mutate(d = Date - age_days) %>%
  ## B
  mutate(s = str_c(year(d) - age_years,"-",month(d),"-",day(d))) %>%
  ## C
  mutate(x = ymd(s)) %>%
  select(Name,x)
```

(A) What do the two rows after comment A do?

(B) After comment B, write down an example of what a string in variable `s` might might look like.

(C) What information is stored in the variable `x`?